

Konzeption, Implementierung und Betrieb eines Volltextservers am Forschungszentrum Jülich

Ulrike Burkard und Dr. Bernhard Mittermaier
u.burkard@fz-juelich.de, b.mittermaier@fz-juelich.de
Forschungszentrum Jülich GmbH, Zentralbibliothek

Zusammenfassung

JUWEL (JUelicher Wissenschaftliche Elektronische Literatur) ist das institutionelle Dokumenten-Repository des Forschungszentrums Jülich, mit dem Ergebnisse Jülicher Forschung einer breiten Öffentlichkeit zugänglich gemacht werden können.

JUWEL bietet den Mitarbeitern des Forschungszentrums die Möglichkeit, Publikationen im Rahmen einer Zweitverwertung (Green Road des Open Access) der Öffentlichkeit kostenlos über das Internet zur Verfügung zu stellen und so eine bessere Verbreitung ihrer Arbeit auch dort zu erreichen, wo ein Zugang zu kostspieligen Zeitschriften und gedruckter Literatur nicht möglich ist.

Im Beitrag wird erst ein kurzer Überblick über das Konzept zu Open Access am Forschungszentrum Jülich gegeben und die Implementierung des Servers am Forschungszentrum erläutert. Anschließend werden die Konzeption, die Struktur und die konkreten Eigenschaften des Servers erläutert.

Jülicher Open Access Modell

Zur Umsetzung von Konzepten von Open Access am Forschungszentrum Jülich wurde das Jülicher Open Access Modell entwickelt. Es soll eine bestmögliche Unterstützung der Autoren im Publikationsprozess erreichen, unabhängig ob der traditionelle Publikationsweg oder eine Veröffentlichung als Open Access Publikation gewünscht wird. Somit wird zum einen Wissenschaftlern geholfen, die ihre Publikationen durch Open Access einer breiteren Öffentlichkeit zugänglich machen möchten, zum anderen bietet dieser Ansatz auch den Wissenschaftlern eine breite Unterstützung, die eine Erstveröffentlichung ihrer Forschungsergebnisse in renommierten Fachzeitschriften anstreben. Auf eine Zwangsverpflichtung für Jülicher Autoren zum Open Access Publizieren wird bewusst verzichtet.

Das Jülicher Open Access Modell verfolgt einen pragmatischen Ansatz und setzt sich aus folgenden drei Komponenten zusammen: dem traditionellen Publikationsprozess, der Unterstützung der Autoren bei der Gestaltung von Autorenverträgen und der Bereitstellung eines Institutionellen Repositoriums:

- **Unterstützung beim etablierten Publikationsprozess:** Viele Autoren möchten den Mehrwert durch den Peer-Review-Prozess und die eingespielten Verbreitungsmöglichkeiten der etablierten Verlage nutzen und mit der Publikation in einer angesehenen Fachzeitschrift hohe Reputation erreichen. Die Zentralbibliothek übernimmt administrative Aufgaben im Kontakt zum Verlag und in der Regel auch die Kosten für Publication Charges und Preprint Gebühren für Angehörige des Forschungszentrums.
- **Unterstützung bei der Gestaltung von Autorenverträgen:** In Zusammenarbeit mit der Rechtsabteilung des Forschungszentrums hat die Zentralbibliothek eine schriftliche Vertragsergänzung für Autorenverträge entwickelt, um den Wissenschaftlern eine Zweitverwertung von Artikeln ggf. auch dann zu ermöglichen, falls eine Zweitverwertung nicht per se erlaubt ist. Diese Klausel wird auf den internen Webseiten der Zentralbibliothek angegeben.
Die Zentralbibliothek prüft auch beim Einstellen von Dokumenten auf den Server

routinemäßig, dass die Copyright-Bestimmungen der einzelnen Verlage auf der Grundlage der SHERPA/RoMEO-Liste¹ erfüllt sind.

- **Dokumentenrepositorium JUWEL:** Mit dem Open Access Server JUWEL werden Publikationen Jülicher Wissenschaftlerinnen und Wissenschaftler langfristig gespeichert, erschlossen und der Öffentlichkeit zugänglich gemacht. Die Zentralbibliothek realisiert die redaktionelle Bearbeitung der Daten und die Systempflege.

Abbildung 1 illustriert die drei Säulen des Jülicher Open Access Modells².

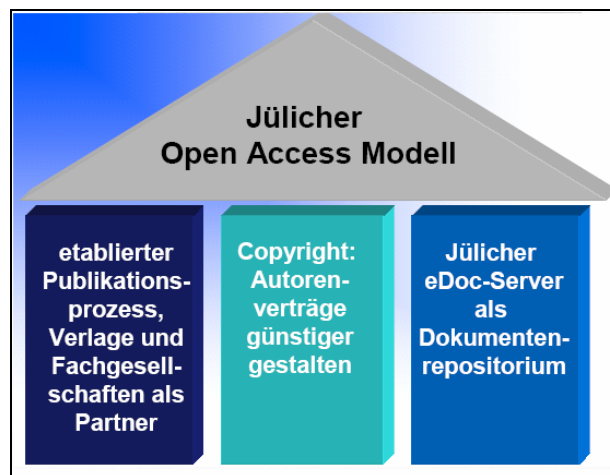


Abbildung 1: Jülicher Open Access Modell

JUWEL als DSpace System

Es gibt eine Vielzahl von Dokumentenserver-Softwaresystemen, viele davon finden eine breite Anwendung und sind teilweise auch als Open Source Software kostenlos zu erhalten. Bei der Auswahl eines geeigneten Systems wurde ein Pflichtenheft für den Server erstellt, ein geeigneter Workflow entworfen und diejenigen Kriterien erarbeitet, die für besonders wichtig erachtet wurden. Neben Zweisprachigkeit (deutsch/englisch) waren dies auch Flexibilität bei Pflichtfeldern, Flexibilität beim Workflow, die Verarbeitung möglichst vieler unterschiedlicher Datentypen sowie eine OAI-PMH-Schnittstelle. Auf der Basis dieser Randbedingungen wurden aus den Angeboten vieler möglicher Systeme drei ausgewählt, die jeweils in fast allen Punkten dem groben Anforderungssystem entsprachen. Diese waren OPUS³, ePrints⁴ und DSpace⁵, die alle in Testinstallationen geprüft wurden. Letztendlich fiel die Wahl auf DSpace, das von der Bibliothek des Massachusetts Institute of Technology (MIT) in Kooperation mit Hewlett-Packard entwickelt wurde und als Open Source Software kostenlos zur Verfügung steht. Weltweit wird es bei über 240 Institutionen als Institutionelles Repositorium eingesetzt.

Als ersten Schritt zur Implementation von DSpace als Dokumentenserver am Forschungszentrum Jülich wurde in Zusammenarbeit mit der UB Dortmund, die schon Erfahrungen mit DSpace ausweisen konnte, eine Testinstallation mit bereits einrichtungsspezifischen Anpassungen eingerichtet. Auf der

¹ <http://www.sherpa.ac.uk/romeo.php> (28.06.2007)

² Ball, R.: "Open Access - Die Revolution im wissenschaftlichen Publizieren?". In: Information zwischen Kultur und Marktwirtschaft : Proceedings des 9. Internationalen Symposiums für Informationswissenschaft (ISI 2004), Chur, 6. - 8. Oktober 2004 / eds.: B. Bekavac, J. Herget, M. Rittberger. - Konstanz, UVK Verlagsgesellschaft, 2004. - (Schriften zur Informationswissenschaft ; 42). - 3-89669-706-4. - S. 413 - 432

³ Werner Stephan und Frank Scholze: "Online Publikationsverbund: Erfassung und Organisation elektronischer Hochschulschriften" In: Bibliotheksdienst Heft 1, 1999, S. 92-102

⁴ <http://www.eprints.org/software/> (28.06.2007)

⁵ <http://www.dspace.org/> (28.06.2007)

Grundlage dieser Installation wurden der Workflow überarbeitet und einige Anpassungen an den systeminternen Vorgängen vorgenommen. Die Installation des eigentlichen Produktionsservers geschah anschließend auf einem Dell PowerEdge 2850 Rechner mit 2 Intel Xeon 3,0 GHz Prozessoren und 4 GByte Arbeitsspeicher. Als Betriebssysteme wird SUSE Linux verwendet. Neben dem Produktionserver wurde ein identischer Spiegelserver mit den gleichen Hardware und Softwaresystemen installiert. Dieser Server dient als Fallback- und Testserver und gibt die Möglichkeit, Softwareupdates oder neue Konfigurationen des Servers an einem Realsystem zu testen, ohne dass es zu Einschränkungen bei der eigentlichen Dienstleistung kommt.

Die eindeutige Identifizierung von Veröffentlichungen in JUWEL erfolgt mittels des Handle-Systems. Diese Entwicklung der *Corporation for National Research Initiatives*⁶ ermöglicht die Vergabe, Administration und Auflösung von Persistent Identifiers in Form von „Handles“⁷.

Struktur und Benutzeroberfläche von JUWEL

JUWEL bietet eine zentral Seite, von der alle Dienste, von der Recherche bis hin zum Einstellen von Dokumenten, erreicht werden können (siehe Abbildung 2).

Die Dokumente in JUWEL sind in sog. Sammlungen organisiert. Jedes Teilinstitut des Forschungszentrums repräsentiert eine Sammlung und auch der Eigenverlag des Forschungszentrums bildet eine eigene Sammlung, in der seine Online-Publikationen verzeichnet sind.

Für die Recherche stehen zwei unterschiedliche Zugänge zur Verfügung: zum einen über Suchfunktionen und zum anderen über Browsingfunktionen. Die einfache Suche durchsucht die Felder Titel, Autor und Schlagwort sowie den Volltext. Die erweiterte Suche ermöglicht eine kombinierte Suche in drei Feldern, wie beispielsweise Autor, Titel, Schlagwort oder Kurzfassung. Mit der Browsing-Funktion kann in den Sammlungen einzelner Institute recherchiert werden sowie im Gesamtindex der Titel oder der Autoren. Auch das Browsen nach dem Erscheinungsdatum ist möglich.



Abbildung 2: JUWEL Startseite

⁶ <http://www.cnri.reston.va.us/> (28.06.2007)

⁷ <http://www.handle.net/> (28.06.2007)

Veröffentlichung auf JUWEL

In JUWEL können Veröffentlichungen aufgenommen werden, falls mindestens einer der Autor Angehöriger des Forschungszentrums Jülich ist und die Arbeit auch im Forschungszentrum Jülich entstanden ist. Gängige Publikationen auf JUWEL sind beispielsweise Zeitschriftenartikel, Reviews, Abstracts, Rezensionen, Reports, Bücher bzw. Beiträge in Büchern oder Proceedings-Bänden, Diplom-Arbeiten, Dissertationen oder Habilitationen. Vereinzelt sind in JUWEL auch Softwareprogramme oder Primärdatensätze verzeichnet. Grundsätzlich ist JUWEL für alle digitalen Dateiformate geeignet, also auch für Bilder, Video- und Audiodateien oder auch für multimediale Lerneinheiten. Bei der Volltextrecherche werden aber nur Textdateien berücksichtigt. Textdokumente werden unabhängig vom Ausgangsformat zusätzlich auch als pdf-Datei abgelegt.

Um eine Aufnahme von Dokumenten in JUWEL zu erreichen muss der Autor den Lizenzbedingungen für die Veröffentlichung auf JUWEL zustimmen. Dabei versichert der Autor, dass mit der Veröffentlichung kein bestehendes Copyright verletzt wird. Unabhängig von dieser Erklärung wird dennoch für jede Veröffentlichung von Seiten der Zentralbibliothek die Copyright-Bestimmung des Verlages auf der Grundlage der SHERPA/RoMEO - Liste geprüft, so dass die Autoren sicher sein können, dass sie nicht mit ihrer Publikation gegen bestehende Copyrightvereinbarungen verstoßen. Mit der Zustimmung zu den Lizenzvereinbarungen räumt der Autor zudem der Zentralbibliothek ein nicht-ausschließliche Recht ein, die Publikation zu speichern, zu vervielfältigen, weltweit zugänglich zu machen und auch bei Bedarf gedruckte und elektronische Kopien anzufertigen. Auch erhält die Zentralbibliothek das Recht, die Ressource bei Bedarf, z.B. im Rahmen einer Langzeitarchivierungsmaßnahme, in andere elektronische und physische Formate zu überführen.

Datenakquise

Eine zentrale Anforderung beim Design von JUWEL war, dass der Aufwand von Seiten der Autoren beim Einstellen der Dokumente möglichst gering gehalten wird. Daher sind neben der manuellen Eingabe der Daten durch den Autor verschiedene weitere Eingabemöglichkeiten ohne größeren Aufwand für die Autoren vorgesehen.

Beim Einspeisen der Daten in JUWEL lassen sich grundsätzlich zwei Ansätze trennen:

1. Daten von zurückliegenden Veröffentlichungen, für die schon Metadaten (und ggf. Volltexte) in einer Datenbank der Zentralbibliothek vorliegen
2. Eingaben neuer Publikationen

Für den ersten Fall werden die schon vorliegenden Metadaten in Form von bibliographischen Angaben und gegebenenfalls auch der Volltext genutzt, um weitgehend automatisiert eine Übertragung der Dokumente in JUWEL zu ermöglichen. Auf diese Art und Weise wurden beispielsweise die Metadaten und Dokumente aus der Datenbank des Eigenverlags des Forschungszentrums ab dem Erscheinungsjahr 2002 automatisiert in JUWEL übertragen.

Ein ähnliches Vorgehen wurde beim Import von sog. Altdaten vorgenommen, d.h. Veröffentlichungen ab 2000, die in Verlagen veröffentlicht wurden, die laut Sherpa/RoMEO-Liste eine Bereitstellung des Original-Verlags-PDFs auf institutionellen Repositorien erlauben. Soweit das Forschungszentrum Jülich entsprechende Zeitschriften abonniert hat, wurden die Artikel manuell über den jeweiligen Zugang heruntergeladen und mit einem Identifikator versehen, durch den die Zuweisung der Metadaten aus der bibliographischen Datenbank des Forschungszentrums und der Transfer der Informationen in JUWEL automatisiert ablaufen konnten.

Im zweiten Fall werden gerade veröffentlichte Publikationen in JUWEL aufgenommen. Dies kann beispielsweise direkt durch den Autor über ein Webinterface geschehen. Der Autor kann sich mit seiner Emailadresse des Forschungszentrums bei JUWEL registrieren lassen und erhält ein Passwort. Nach einer Identifizierung mit diesen persönlichen Angaben erhält der Autor dann die Möglichkeit für

die Sammlung seines Instituts Publikationen in JUWEL einzutragen: In einem mehrschrittigen Prozess werden dabei zuerst die bibliographischen Metadaten der Publikation aufgenommen. Die sachliche Erschließung geschieht anschließend zum einen durch Auswahl von Sachgruppen aus einer vorgegebenen Liste und zum anderen durch die Angabe von Schlagworten. Zudem wird ein zusätzliches Feld angeboten, in das der Abstract der Veröffentlichung eingefügt werden kann. Schließlich kann der Volltext hochgeladen werden. Die Eingabe der Metadaten wird mit einer Übersichtsseite abgeschlossen, die die Metadaten des Dokumentes nochmals anzeigt und Möglichkeiten für Korrekturen bietet.

Der letzte Schritt beim Einstellen eines Dokumentes ist die Anzeige der Lizenzbedingungen, die der Autor für die endgültige Freischaltung seines Dokumentes in JUWEL akzeptieren muss.

Eine weitere Möglichkeit, eine Publikation in JUWEL aufnehmen zu lassen, ist das Senden einer Emailnachricht an eine Funktionsadresse der Zentralbibliothek. Hierbei geschieht die manuelle Eingabe der Metadaten und der Upload des Volltextes durch Mitarbeiter der Zentralbibliothek.

Am effizientesten geschieht die Aufnahme von Publikationen in JUWEL jedoch über die Veröffentlichungsdatenbank, eine zentrale bibliographische Datenbank, die alle Veröffentlichungen von Mitarbeitern des Forschungszentrums verzeichnet. Für diese Datenbank wurde eine Möglichkeit geschaffen, bei der Eingabe der Metadaten, die für jede Publikation gemeldet werden müssen, auch Volltexte hochzuladen. In periodischen Abständen werden diese Dokumente und die dazugehörigen Metadaten automatisiert in JUWEL überführt.

Abbildung 3 fasst die verschiedenen Arten der Datenakquise für JUWEL zusammen:

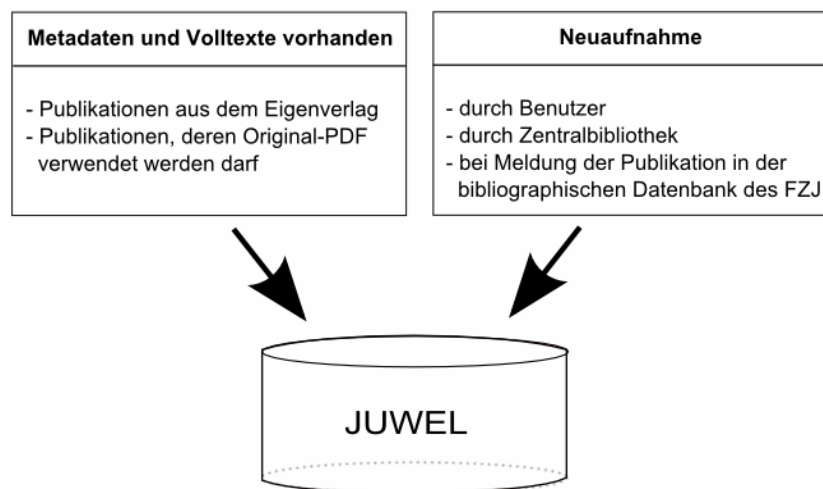


Abbildung 3: Datenakquise in JUWEL

Bestandsentwicklung von JUWEL

Nach der endgültigen Installation und Konfiguration von JUWEL wurden als erste Dokumente die Online-Publikationen des Eigenverlages in JUWEL eingespeist, so dass zum Zeitpunkt der Eröffnung von JUWEL Anfang Juli 2006 schon ca. 390 Dokumente in JUWEL zur Verfügung standen. Mit der Freischaltung des Servers wurde eine Roadshow innerhalb des Forschungszentrums gestartet, bei der JUWEL in den Instituten vorgestellt wurde und den Wissenschaftlern die Vorteile einer Zweitveröffentlichung auf JUWEL aufgezeigt wurden.

In den ersten Betriebsmonaten wurden von Seiten der Autoren nur sehr zögerlich Dokumente eingestellt, auch das Angebot der Eingabe durch die Zentralbibliothek wurde nur zurückhaltend genutzt. Mittlerweile haben jedoch mehr und mehr Wissenschaftler den Mehrwert des Servers als

weitere Verbreitungsmöglichkeit ihrer Publikationen erkannt, was sich in einer stetigen Zunahme der an JUWEL gelieferten Dokumente zeigt. Seit September 2006 ist auch die Eingabe von Volltexten bei der Meldung von Veröffentlichungen in die Veröffentlichungsdatenbank (s.o.) möglich. Auch dies hat zu einer weiteren Steigerung der Anzahl der Dokumente beigetragen.

Den größten Beitrag zur Entwicklung des Bestandes in JUWEL hat jedoch der Import von Altdaten geleistet: Veröffentlichungen aus dem Forschungszentrum Jülich, die nach 1999 erscheinen sind und bei denen laut Lizenzeintrag bei Sherpa/RoMEO eine Verwendung von Original-PDFs auf dem Dokumentenserver erlaubt ist, wurden in JUWEL aufgenommen. Mit dem Stand vom 28.06.2007, also nach ca. einem Jahr Laufzeit, weist JUWEL nun 2832 Dokumente auf.

Nutzung von JUWEL

Um weiterhin die Nutzung von JUWEL, sowohl von Seiten der Autoren als auch von Seiten externer Nutzer weiter zu steigern werden fortlaufend weitere Werbemaßnahmen unternommen. Dabei geschieht Werbung für JUWEL innerhalb und außerhalb des Forschungszentrums primär auf zwei Wegen: Zum einen werden regelmäßige Präsentationen durchgeführt, sei es im Rahmen von Schulungsangeboten der Zentralbibliothek oder direkt vor Ort in Veranstaltungen der jeweiligen Institute bzw. Arbeitsgruppen. Zum anderen wurden zwei Flyer entworfen, die den Bekanntheitsgrad von JUWEL innerhalb und außerhalb des Forschungszentrums Jülich weiter erhöhen sollen: ein interner Flyer für Wissenschaftler des Forschungszentrums stellt die Möglichkeiten von JUWEL als institutionelles Repositorium vor und gibt Informationen zu den verschiedenen Arten der Eingabe von Dokumenten. Der zweite Flyer ist für Personen außerhalb des Forschungszentrums gedacht. Er gibt einen Überblick über die Aktivitäten der Zentralbibliothek im Bereich Open Access und stellt JUWEL aus der Sicht eines Nutzers dar, der nach Publikationen des Forschungszentrums Jülich recherchiert.

Mit der Freischaltung wurde JUWEL bei verschiedenen Harvestern, wie beispielsweise OAster⁸, BASE⁹ oder ScientificCommons¹⁰ angemeldet. Diese zentralen Archive fragen die Daten aus den verschiedensten Repositorien ab und bieten sie zusammengefasst als Sammlung an und ermöglichen eine zentrale Suche in allen angeschlossenen Repositorien.

Auch Suchmaschinen wie Google¹¹ haben die Inhalte von JUWEL indiziert. Die Suche nach dem Titel einer in JUWEL gespeicherten Veröffentlichung enthält in der Regel unter den drei ersten Suchergebnissen den URI bei JUWEL.

Die Nutzungszahlen von JUWEL sind seit der Inbetriebnahme von JUWEL stetig gewachsen. Abbildung 4 zeigt die Entwicklung der Zugriffe seit der Inbetriebnahme von JUWEL vor ca. einem Jahr an.

⁸ <http://www.oaister.org/> (28.06.2007)

⁹ <http://base.ub.uni-bielefeld.de/index.html> (28.06.2007)

¹⁰ <http://de.scientificcommons.org/> (28.06.2007)

¹¹ <http://www.google.de> (28.06.2007)

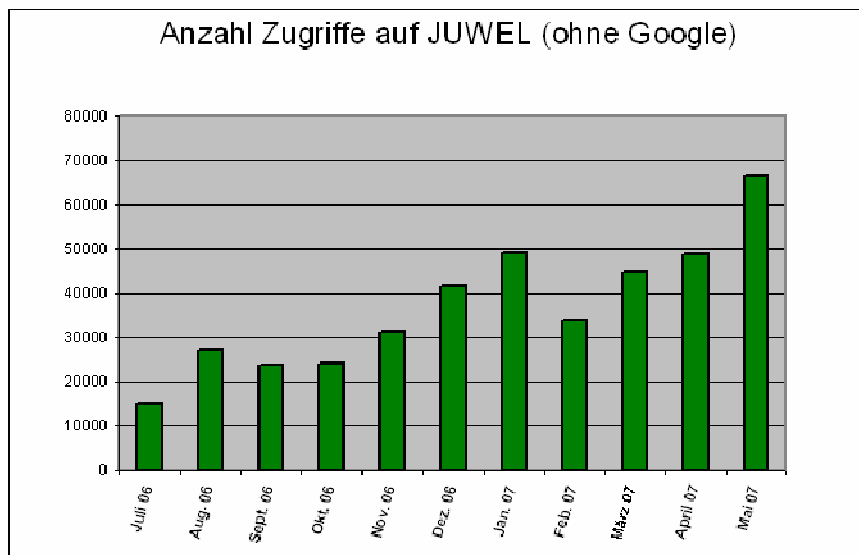


Abbildung 4: Anzahl der Zugriffe auf JUWEL ohne Google-Zugriffe

Die stärksten Zugriffe weisen erwartungsgemäß Suchmaschinen wie Google auf (über 1 Million im ersten Betriebsjahr), aber auch die Nutzung von anderen Seiten hat mit über 66.000 im Monat Mai einen neuen Höchststand erreicht.